Research Article

Analysis of Rural Tourism Demand Characteristics and Experience Differences Based on Association Rule Mining

Qun Jiang

Wuhan Polytechnic, Wuhan, Hubei 430074, China

Correspondence should be addressed to Qun Jiang; m15038169762@126.com

Received 25 September 2021; Revised 10 October 2021; Accepted 12 October 2021; Published 13 November 2021

Academic Editor: Xin Ning

Copyright © 2021 Qun Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Whether the characteristics of rural tourism changes or not provides the scale and basis for judging whether the rural tourism landscape has changed, but it cannot provide a judgment on the impact of rural tourism landscape changes. The impact is relative to the rural tourism landscape goal. The determination of rural tourism landscape objectives provides a baseline for judging the direction and impact of rural tourism characteristics and provides a prerequisite for rural tourism landscape actions. The determination of the quality target of the rural tourism landscape is mainly determined by the internal process and external demand of the rural tourism landscape. Through in-depth research on the frequent pattern growth algorithm FP-Growth, the algorithm can find frequent item sets by not generating candidate item sets. The core of the algorithm is the frequent pattern tree FP-tree, which can efficiently compress the transaction database. Based on the advantages of FP-tree, this paper improves a FP_Apriori algorithm based on frequent pattern trees. This algorithm projects the entire transaction database onto the FP-tree, avoiding a lot of I/O overhead. At the same time, I propose a more directional and targeted search strategy for FP-tree, which reduces the running time of the algorithm and uses the principle of the Mapping_Apriori algorithm to prethin the frequent item sets. This article uses the text analysis method of network data to excavate the characteristics and internal structure of rural tourism demand. The rural tourism market has a wide range of needs and multiple levels, and traditional research methods such as questionnaires have limited sample size and sample structure. With the help of network data, text mining, and other statistical analysis methods, in-depth empirical research on the characteristics and spatial structure of rural tourism in a certain region can cover more research groups. The research confirms that the results of using text analysis and questionnaire analysis on the perception of destination image are relatively consistent. Therefore, the network text analysis method is an effective tool to study the demand structure of the rural tourism market.

1. Introduction

Data mining is to dig out hidden, unknown relationships, patterns, or trends that have potential value to decision makers from large-scale heterogeneous or multisource data and use these knowledge and laws to establish auxiliary decision-making or prediction models [1]. It is the process of using various analysis tools to discover models or potential relationships among massive amounts of data. Because data mining can find useful information in data and make enterprises profitable, this makes the application of data mining technology more and more common. The commonly used methods for data analysis using data mining

mainly include classification, regression analysis, clustering, association rules, time series patterns, and deviation analysis, which mine data from different perspectives [2]. Among them, the association rule technology is a classic data mining method, which can discover the direct potential interrelationships of data, and this relationship does not need to be directly expressed in the database [3]. The task of association analysis is to discover the degree of association or association rules between things [4].

In recent years, some places in China have recognized the development law of local rural tourism [5]. After fully understanding the spatial laws of their rural tourism, cities have replanned, adjusted, and optimized the spatial pattern

of rural tourism and proposed strategies or countermeasures for the development of rural tourism space according to local conditions [6]. Studying the spatial change of rural tourism is helpful to discover the process of agglomeration and diffusion of tourism economic activities in the region; find the core node, development axis, or development zone; and establish a regional tourism space that can integrate social, cultural, economic, and environmental resources. By constructing rural tourism centers of different levels and levels based on the centrality of the rural tourism spatial network and formulating corresponding development strategies, it is possible to coordinate the rational development of regional rural tourism and integrate tourism, leisure, specialty food production and consumption, and e-commerce [7]. Studying the spatial changes of rural tourism is helpful to develop and construct new tourism infrastructure and to selectively retain existing facilities to realize scientific rural tourism spatial planning [8]. Whether it is the transformation and construction of rural tourism consulting centers, tourism distribution centers, or rural tourism landscape tourism villages with rural characteristics, the spatial layout and planning of rural tourism are always the constraints of the construction of these centers and characteristic villages and towns, and they are the effective use of different types of centers [9].

Aiming at the improvement of the Apriori algorithm, this paper mainly proposes the FP_Apriori algorithm based on the frequent pattern tree. The improvement ideas of the algorithm are mainly reflected in three aspects. First, in view of the shortcomings of traditional algorithms that scan the transaction database multiple times, the frequent pattern tree is used to store transactions, so that only one traversal scan of the transaction database is required. Second, I optimize the complexity coefficient of the frequency calculation of the candidate item set to improve the running speed and mining efficiency of the algorithm. Third, I take advantage of the nature of frequent item sets to slim down L_{k-1} and remove some useless candidate item sets in advance. In addition, this paper makes a very detailed research and analysis on the improvement ideas of the FP_Apriori algorithm, analyzes the optimization strategy of FP-tree and candidate set frequency calculation, and analyzes the core steps of the algorithm. With the help of more than 20,000 online review data on 5 online platforms from 50 major rural tourist attractions in a certain region, this paper explores the specific research content of the seasonal cycle, demand attributes, supply characteristics, and supply-demand gap of rural tourism market in a certain region. The empirical research reveals that the demand enthusiasm of the rural tourism market in a certain region has obvious dual-peak characteristics, insufficient demand upgrades, and the homogenization and simplification of tourism supply content. The overall satisfaction of the market is relatively good, the demand loss trend is obvious.

2. Related Work

A key problem in association rule mining is to find all frequent item sets that meet the minimum support given by the user from the transaction set [10]. This step concentrates all the calculations. The original algorithm to solve the association rule problem is the AIS algorithm. In order to

improve the AIS algorithm, related scholars have proposed the OCD algorithm [11]. The OCD algorithm uses the combined information of the previous search to reduce the number of candidate item sets produced this time. Later, researchers proposed the most famous Apriori algorithm in association rule mining and its variants AprioriTid and AprioriHybrid algorithms to discover frequent item sets [12]. Since then, many scholars have proposed algorithms for discovering frequent item sets of association rules, but most of the algorithms are variants or improvements of the Apriori algorithm [13, 14]. Since the Apriori algorithm is a multipass search algorithm, for a large data collection, the external memory must be read once for each search, and the I/O overhead is very high [15]. Therefore, most of the improved algorithms are making a fuss on how to reduce the number of searches. In fact, what really affects the efficiency of the classic association rule frequent item set discovery algorithm based on the Apriori algorithm is the calculation of the item set and its support. If the number of different items included in the transaction data set is n , the frequent item set discovery algorithm based on Apriori will calculate $2n$ item sets. When n is relatively large, a combinatorial explosion will occur. In fact, this will be an NP-hard problem.

The sampling algorithm randomly samples a part of the sample from the original data set and uses the sample to mine association rules to reduce the number of searches of the algorithm. However, because the data set often has uneven data distribution, random sampling cannot guarantee that it can be drawn at all. Although the Partition algorithm reduces the burden of I/O by separately mining the data collection partitions and finally summarizes it, it actually increases the burden of the CPU. The DIC algorithm uses a dynamic calculation strategy to reduce the number of searches to improve the efficiency of the algorithm, but there is no fundamental difference in thinking from the Apriori algorithm, and it is also a multi-pass search algorithm. These algorithms generate candidate item sets when reading in transaction data, and generate many unnecessary candidate item sets, which are computationally intensive. Especially for massive data sets, the above algorithms can only have a certain mining efficiency under high minimum support and minimum credibility or after other constraints are added. Otherwise, a combination explosion of frequent item sets will occur. The inefficiency even exceeds the storage and computing power of the machine. Because any algorithm must calculate the item set and its support, what really affects the efficiency of the algorithm is the calculation of the item set and its support. Each calculation not only takes a lot of CPU time but also involves I/O requests. Therefore, only by essentially reducing the generation of unnecessary item sets and reducing the calculation time for the support of item sets can the number of frequent item set generation and data mining time be greatly reduced.

The academic circle mainly studies the spatial distribution characteristics of rural tourism from the aspects of integrated destination management, location advantages, and resource environment [16]. For the integrated management of tourist destinations and the optimal allocation of resources, related scholars have studied the central structure of Korean rural

space and its corresponding development strategies based on the resources of rural convenience facilities [17]. Advantageous locations or traffic arteries have led to changes in the spatial pattern of rural tourism. Researchers reviewed tourism development studies and found that rural areas on the fringe of cities strongly attract day-trip tourists, and the surrounding areas attract low-level tourists [18]. Relevant scholars have studied the relationship between high-speed rail and tourist attraction between Perpignan, France, and Barcelona, Spain, and found that high-speed rail reduces transportation costs, improves the accessibility of destinations, and enhances the spatial competitiveness between destinations [19]. It promotes the concentration of tourism activities in Barcelona, but it is detrimental to Perpignan. In addition, some scholars have studied the spatial relationship between the comfortable resources of cultural facilities in Korea, the spatial relationship between the hot tub in the Appalachian Mountains, Ohio, and the suitability of forest tourism land and found out how to influence the superior tourism resources or superior environment in the region [20]. The tourism policy and development planning strategy of the development department will affect the spatial layout of rural tourism, and so on. Research on the spatial distribution patterns of foreign scholars does not directly start from the spatial distribution or structure of rural tourism resources or scenic spots but indirect research from the hierarchical system, attractive conditions, development conditions, and population distribution of rural tourism resources or scenic spots [21].

Changes in rural tourism have an impact on the development positioning, industrial structure, and employment of the entire tourism industry. Research believes that tourism is a policy choice of local governments and a more general partial response of rural areas to the ever-changing continental and global economy [22]. As a new economic activity that replaces agriculture, tourism has changed the agricultural structure, restructured the countryside, and changed government decisions and long-term family decisions, directly affecting the organizational structure of rural communities, environmental quality, and inter-city relations. In fact, the integration effect of rural space is very obvious. Rural space is no longer purely an area related to the production of agricultural products but is regarded as a place to stimulate new social and economic activities. It usually integrates tourism, leisure, and specialty food. The large-scale development of tourism in the periphery of remote villages can promote the transfer of remote rural labor force and create tourism economic links with poor families. However, it is worth noting that tourism is a low-risk, low-return livelihood strategy. It can be seen that the development of rural tourism is not only a policy choice for the government in the face of global restructuring and local response but also a way of economic activity and employment problems.

3. The Characteristics of Rural Tourism as a Spatial Analysis to Manage the Change of Rural Tourism Landscape Experience

3.1. The Overall Framework of the Classification of Rural Tourism Characteristics. In order to continuously promote the understanding of rural tourism landscape and the man-

agement of rural tourism landscape, it is very important to establish a frame of reference that can communicate with rural tourism landscape. However, the classification process of “rural tourism landscape” is very complicated, because this object contains multiple dimensions, as well as human perception and physical reality. The rural tourism landscape is composed of components that appear in the “view field,” including landforms, water bodies, vegetation, and infrastructure. These are often referred to as rural tourism landscape layers. Although it is very important to decompose the rural tourism landscape into different layers, due to the nature of the integrity and complexity of the rural tourism landscape, the interaction of these layers forms the rural tourism landscape. The rural tourism landscape as a whole operates under a nested scale and needs to be linked to the administrative management level. The DPSIR framework of rural tourism landscape changes is shown in Figure 1.

The classification system of rural tourism characteristics is a parallel structural relationship, which is a flexible framework that can be continuously expanded. The dimensional division is the choice of perspective for observing, understanding, and regulating the object of the rural tourism landscape. The selection of dimensionality is closely related to the goal, level, and scale. The more dimensions are divided, the more thorough and comprehensive the understanding of rural tourism landscape. Therefore, the classification system of rural tourism characteristics can be applied to different spatial structure levels and administrative management systems. Whether it is a transnational scale, a national or regional scale, or a local scale, a hierarchical nesting relationship is formed between different scales.

From the perspective of linguistics, corresponding to the planning process, the classification of rural tourism characteristics as a specific planning language also has three purposes or three modes. The first is the instruction model—the intuitive phenomenon that describes the characteristics of rural tourism; the second is the evaluation model that expresses subjective value judgments—the evaluation of the value and quality of the characteristics of rural tourism; the third is the prescribed mode that requires others to act—regulate the future villages’ goals and methods of tourism characteristics. There is a progressive relationship among the three models, but they should be based on common problems or around common goals, such as the alienation and decline of rural and rural tourism landscapes. First, the characteristics of rural and rural tourism need to be defined, analyzed, and described. And then, you make corresponding evaluations on the quality and value of rural tourism characteristics from the perspective of human use and environmental impact. Finally, based on the goals of sustainable development, the quality objectives of rural tourism landscapes are stipulated, so as to form corresponding guiding strategies (protection, management, or planning) to protect, strengthen, or restore the characteristics of rural tourism. The basic purpose of the classification of rural tourism characteristics is to serve the planning and management of the rural tourism landscape.

3.2. Evaluation and Decision-Making of Rural Tourism Characteristics. The evaluation stage of rural tourism

FIGURE 2: Process diagram of rural tourism feature evaluation.

problems and the goals that must be met and the objects of use. The purpose of the evaluation is to ensure that the changes in land use and the planning and design of development projects achieve a harmonious relationship with the surrounding environment, or to further strengthen or shape new rural tourism landscapes. The evaluation method should be as objective as possible, and establish clear logic and ensure its transparency. Inevitable subjective evaluation (such as the value of rural tourism landscape) requires the participation of professionals, and should be combined with the historical characteristics of the rural tourism landscape. This process should ensure the participation of stakeholders as much as possible, and the extent and nature of their participation should be clarified. The process of rural tourism feature evaluation is shown in Figure 2.

Based on the analysis and evaluation of rural tourism characteristics, the evaluation conclusions can be drawn to assist decision-making, including the determination of rural tourism landscape quality objectives for each rural tourism landscape type or region and corresponding rural tourism landscape actions. The type of evaluation conclusion is also closely related to the specific user population.

In the decision-making stage of the evaluation of rural tourism characteristics, based on the results of investigation, analysis, and evaluation, the subjective judgment of the researcher, the application method, and the opinions of stakeholders are added, that is, “who” will implement determines the development of the decision-making stage. This process will inevitably have various subjectivities, so a clear process and causality are very important. In the process of evaluation formation and the use of evaluation results, it is necessary to understand which elements are relatively objective and noncontroversial and which elements are prone to produce different opinions. Therefore, in the whole process,

the participation of stakeholders and local people is also necessary and very important.

3.3. Goals and Decision-Making Forms of Rural Tourism Landscape Governance. The protection, restoration, and strengthening of rural tourism characteristics are not intended to be an obstacle to the creation of new rural tourism landscapes. When it is concluded that certain types of rural tourism characteristics or areas are suitable for strengthening or renewal strategies based on evaluation and analysis, this means that there is room for huge changes in such rural tourism landscapes. Similarly, many degraded and declining rural tourism landscapes also urgently need active rural tourism landscape reconstruction to improve environmental quality and people’s quality of life, such as old industrial areas, wetlands, and swamp areas that need to be rebuilt and renewed. The analysis of rural tourism characteristics plays a key role in determining the rural tourism landscape areas with the potential for renewal and promotion. They can assist the restoration of the lost valuable rural tourism features and the investigation of the creation of new rural tourism landscapes.

The method of making judgments based on characteristics will vary with the transaction to be determined and the specific situation that must be implemented. There must be a clear logic behind the method of goal judgment. It is necessary to carefully consider the overall characteristics and key characteristics of the rural tourism landscape, its history and origin, recent changes, changing trends, and future driving forces. There may be the creation of new rural tourism. This will help realize rational judgments about future changes in existing features.

Development control also needs to integrate the characteristics of rural tourism. The purpose is to ensure that the

decision-making can combine the characteristics of rural tourism as much as possible. The feasibility of the development project mainly depends on whether the development behavior will have a negative impact on the rural tourism landscape in the area. When relevant agencies adopt this method, they must formulate effective evaluation criteria (based on the key rural tourism characteristics and the impact analysis of the development-sensitive rural tourism characteristics) to judge it. Development must meet the characteristics of the rural tourism policy objectives of the region, the developer knows how to obtain the permit, and the personnel of the relevant agencies know how to review the development plan and finally form a high-quality development and construction.

The method of rural tourism landscape strategy is mainly aimed at how the rural tourism landscape changes, using the assessment of rural tourism characteristics as the spatial framework to determine the development strategies (strengthening, maintaining, restoring, etc.) of different regions. The coordination and judgment processes supporting the formation of these strategies need to be clear and transparent, usually involving subjects with different values and different rural tourism landscape objectives, and are closely related to the rural tourism landscape quality objectives. For example, in LEED's rural tourism landscape strategy, different development strategies are determined based on the characteristic areas of rural tourism, such as protection, restoration or strengthening strategies. It can be seen that different from the single protection method in the past, the document provides more flexible strategies to guide different Regional development and changes. In the rural tourism landscape policy division of Staffordshire, the development strategy is further upgraded to a policy division. The policy objectives include five categories: innovative renewal, restoration, strengthening, maintenance, active protection, and areas characterized by rural tourism. The smallest unit of "Land Description Unit" (LDU) is the spatial framework to implement various policy objectives into specific spatial scope, as the basis for future judgment and decision-making.

Environmental impact assessment or strategic environmental assessment for different types of development is an important part of the development and construction assessment of the European Union and England. Rural tourism characteristics and visual impact assessment are the main components of the research. The rural tourism landscape department and the environmental assessment department finally form an independent report. The general principles and core of the method are similar in each case. The assessment of rural tourism landscape impact should define the type of rural tourism landscape in the area where the development project is located. It analyzes how the various elements that make up the rural tourism landscape, the aesthetic and perception aspects, the characteristics, and the key characteristics that form the characteristics are affected. It assesses the consequences (profits and losses) of not carrying out development and construction and assesses the measures and scope of mitigating the impact. Rural tourism feature assessment can guide all stages of development and construction under the framework of environmental

impact assessment, so as to guide construction projects to respond better to the background of rural tourism landscape.

4. FP_Apriori Algorithm Based on Frequent Pattern Tree FP-Tree

4.1. The Idea of FP_Apriori Algorithm. Through the analysis of the FP-Growth algorithm, a large transaction library can be efficiently compressed using the frequent pattern tree FP-tree, which can completely avoid the huge overhead of repeatedly scanning the transaction library, which directly and effectively solves the classic Apriori algorithm relying heavily on the I/O overhead of the system.

In the traditional Apriori algorithm, we use L_{k-1} self-connection to get the candidate item set C_k and then scan the database to get the frequent item set L_k . In this process, there are two major shortcomings: the candidate item set C_k is too bloated, which will produce a particularly large number of useless candidate item sets; second, when determining the frequency of the candidate item set, the entire database needs to be repeatedly traversed and scanned. The operating efficiency is very poor. We can also thin the frequent item set L_{k-1} in advance according to the pruning strategy, and optimize the frequency calculation process of the generated candidate item set. The specific optimization strategy will be studied in detail below.

Combining the above improvement ideas can effectively solve and overcome the shortcomings of the traditional Apriori algorithm, thereby further improving the execution efficiency and applicability of the algorithm, and can also serve as a reference for the improvement of other association rule algorithms.

4.2. Frequent Pattern Tree FP-Tree. First, I give the relevant definition of the frequent pattern tree FP-tree.

Definition 1. Frequent pattern tree FP-tree is composed of a root node (null), a group of root node item prefix subtrees, and an item header table.

Definition 2. All nodes of the item prefix subtree are composed of three parts: item name, node count, and node chain. Among them, the count of the node represents the number of transactions that contain the node in the database, and the node chain points to the next node with the same item name in the frequent pattern tree. If there is no next node, it is empty. In particular, in the FP_Apriori algorithm, some improvements need to be made to FP-tree; that is, all nodes of the item prefix subtree also need to record the name of the transaction that contains the node.

Definition 3. The item header table mainly contains two attributes: the item name and the node chain head. The node chain head is the node that points to the first item in the FP-tree.

I generate a bit matrix containing only 0 and 1 by scanning the transaction library. In this matrix, each row represents a transaction T_i , and each column represents an item

set I_j , where 0 and 1, respectively, indicate whether the item set I_j is in appears in this transaction T_i . Suppose the transaction set is

$$T = [T_0 \ T_1 \ T_2 \ \cdots \ T_{m-2} \ T_{m-1}]. \quad (1)$$

The collection of all item sets is

$$I = [I_0 \ I_1 \ I_2 \ \cdots \ I_{n-2} \ I_{n-1}]. \quad (2)$$

The specific definition of the matrix is

$$M = (T_{ij})_{(m-1) \times (n-1)} \\ = \begin{cases} T_{ij} = -1, & I_j \notin T_i \\ T_{ij} = 1, & I_j \in T_i, \end{cases} \quad i = 0, 1, 2, \dots, m-1, j = 0, 1, 2, \dots, n-1. \quad (3)$$

Insert a column into the matrix M to record the number of transactions with the same item set, denoted as RC; if there are no duplicate transactions in the transaction database, set this value to 1. In this way, the order of the matrix is reduced very effectively. The specific definition of RC is

$$RC_i = \begin{cases} RC_i - 1, & T_{sj} = T_{ij}, \\ -1, & T_{sj} \neq T_{ij}, \end{cases} \quad j = 0, 1, 2, \dots, n-1, \quad t = 0, 1, 2, \dots, m-1. \quad (4)$$

The support of all 1st-order item sets is the sum of non-zero elements in each column, which is specifically defined as

$$\text{Sum} = \sum_{i=0}^{m-1} T_{ij} T_{ij} > 0, \quad j = 0, 1, 2, \dots, n-1. \quad (5)$$

The support of all k -th order item sets can be calculated by RC and T_{ij} , which is specifically defined as

$$\text{Count}(k\text{-Item set}) = \sum_{i=1}^n \sum_{j=1}^k (T_{ij} \cdot RC_i). \quad (6)$$

According to the above definition, the specific algorithm for constructing the FP-tree can be given as shown in Figure 3.

4.3. Optimization Strategy for Frequency Calculation of Candidate Item Sets. In the improved algorithm FP_Apriori in this article, first, you scan the entire transaction database to get the first-order frequent item set L_1 and its corresponding support and transaction identification T_{id} and then continuously iterate through L_1 to generate $L_2, L_3, \dots, L_k$.

After self-connection through L_1 , the second-order candidate set $C_2(x, y)$ is obtained, where x and y are both items in the second-order candidate set C_2 . At this time, before searching the frequent pattern tree FP-tree to calculate the candidate item set frequency, I obtain the item with less sup-

port between x and y according to L_1 . In this way, calculating the candidate item set frequency only needs to scan the transaction database. Specific transactions do not need to scan the entire database; in other words, we only need to search for specific branches of the FP-tree to calculate the frequency of candidate item sets. Similarly, I generate $C_3(x, y, z)$ through L_2 self-connection; first, we find the item with the least support of x, y , and z . At this time, you only need to search for the specific branch of the FP-tree to calculate the $C_3(x, y, z)$ support degree.

Through the above optimization strategy, the support degree of the candidate item set can be calculated faster and more directly, and to a certain extent, the execution efficiency of the algorithm can be greatly improved.

4.4. FP_Apriori Algorithm Analysis. The FP_Apriori algorithm is a comprehensive manifestation of the traditional Apriori algorithm and the frequent pattern tree FP-tree. Transplanting FP-tree to the Apriori algorithm can effectively improve the execution efficiency of the algorithm. The specific improvement work of FP_Apriori is mainly reflected as follows:

- (1) The FP_Apriori algorithm uses FP-tree to compress the transaction database, which can specifically solve the problems of the traditional algorithm scanning the database too many times, the I/O burden being too heavy, and the efficiency being too low
- (2) It adopts the strategy of prethinning L_{k-1} by using the Mapping_Apriori algorithm. The C_k generated by the self-connection of L_{k-1} of this strategy is a necessary condition for frequent item sets. The frequency of the first i items in any item set of L_{k-1} is at least $k-i$. In the same way, the FP_Apriori algorithm can also thin L_1 in advance to avoid the generation of useless candidate items

5. Experiment and Analysis

5.1. Data Source. This paper collects network comment data of 50 rural tourist spots as basic research data. Considering that the rural tourism market mainly comes from the local area, and the O2O platform for local life is also an important channel for residents to purchase and evaluate leisure experience products, this study added Dianping.com as a network data source website. Dianping.com includes small- and medium-sized rural tourist spots that are not covered by other large-scale tourism websites. The emphasis of the word-of-mouth content of each website is different, which can ensure the richness and completeness of the information source. The information comparison of the five text source websites is shown in Figure 4.

5.2. Data Characteristics. I read through all the samples to screen the samples one by one and delete duplicate and irrelevant information. The word-of-mouth evaluation needs to include the comprehensive experience evaluation of the project; that is, it needs to include the tourist's travel itinerary in rural tourism, the same tourist group, service quality,

FIGURE 3: Algorithm for generating frequent pattern tree FP-tree.

FIGURE 4: Comparison of text source website information.

environmental quality, accommodation conditions, transportation accessibility, product evaluation, willingness to revisit, etc. After screening and cleaning, there are 19,916 remaining. Flower port and urban vegetable garden in a certain area are national 4A-level scenic spots. The distribution of online text data among scenic spots indirectly supports

the problem of market concentration in the rural tourism industry, that is, the high market concentration of farm products. Although there are a large number of rural tourism projects in a certain area, the well-known farms (villas) are the main force for receiving rural tourists. The number of reviews in each month is shown in Figure 5.

FIGURE 5: Distribution of the number of text data before filtering.

FIGURE 6: Monthly distribution of rural tourism tourist volume data.

5.3. Analysis of Market Demand Heat. Network text analysis methods are used more and more widely in tourism research, especially the research on the image of tourist destinations and the experience of tourists. Collecting online travel word-of-mouth from multiple destinations for text analysis and cross-comparative analysis can not only refine the market demand characteristics of tourists but also study the supply differences between destinations and find the gap between destination supply and demand. This research will use about 20,000 online review data from 50 major rural tourist spots in a certain area on 5 major online platforms and use text analysis methods, statistical testing methods, cluster analysis methods, common network analysis, and

TABLE 1: Test of the relationship between demand heat and whether the scenic spot has won the title.

t	Df	Sig.	t test for the mean equation			
			Mean difference	Standard error	95% lower limit	95% upper limit
1.66	43	0.08	28	18	2	64
2.21	46.3	0.03	28	13	5	55

other statistical methods for analysis. We combine the research on the characteristics of rural tourism demand in a certain area with the characteristics of tourism supply of rural tourism projects to find the direction of the

FIGURE 7: Common feature extraction results of rural tourism experience.

optimization of the supply and demand structure of rural tourism in a certain area.

Although there is a time lag between the travel time and the review time of tourists, the time distribution of the network review data of scenic spots can basically reflect the travel situation of the market in different time periods. The number of online reviews of major rural tourism projects in a certain area has two obvious peak periods, and there are differences in geographical space, reflecting the seasonal cycle of the industry. As shown in Figure 6, September of the year is the peak periods for online reviews. After calculation, the monthly average value of the number of reviews in scenic spots is 1,660, and the coefficient of variation is 0.204. The number of reviews in September has greatly exceeded the average. Based on the specific comments, it can be seen that there is a concentrated period of outings for residents in a certain area, flower viewing during festivals, and citizens going to the suburbs to sweep graves. The number of reviews in the remaining months is more evenly distributed, which is the result of multiple factors such as the repetitiveness of rural tourism demand, the convenience of urban transportation, and the convenience of online platforms. A stable passenger flow helps to maintain the balance of rural tourism supply.

As branded rural tourism projects have a high degree of popularity, I will further examine the relationship between the popularity of reviews and whether scenic spots have received relevant titles, that is, the relationship between the popularity of scenic spots and whether they are scenic spots above 3A or demonstration sites for leisure agriculture and rural tourism. I mark the scenic spots that belong to at least one of these types as “1,” and mark them as “0” if they are neither, and perform an independent sample t test. The test results are shown in Table 1.

Levene’s test confirms that the variances are homogeneous, and “assuming the variances are equal” is used. The

P value of the two-tailed significance level of the independent sample t test is 0.094, which is significant at the 10% level. The test results show that the average monthly online reviews of scenic spots above 3A and demonstration sites for leisure agriculture and rural tourism are higher than those of rural tourism projects that have not obtained the above titles. These projects have a relatively high level of reception, which also confirms the market utility and brand utility of various standardized demonstration sites.

5.4. Demand Experience Feature Extraction Experiment. By extracting the basic commonality of tourism demand and experience of each project, the demand attribute of rural tourism industry is analyzed. The word-of-mouth documents of each scenic spot are processed one by one, and the software is used to extract the high-frequency feature words of the text documents of each scenic spot. In order to ensure the completeness of the travel experience summary and the accuracy of the description, according to the attribute classification of nouns, verbs, and adjectives, the top ten words of the three types of words in the document are extracted, totaling 30 words. I check the high-frequency feature words extracted from the text research files of each scenic spot one by one, remove words that do not contain tourism experience and other tourism characterization meanings, remove the vocabulary that characterizes the scenic spot name and the area, and supplement it with subsequent vocabulary. According to the frequency of the vocabulary of all 30 high-frequency words in each scenic spot, they are extracted at least sequentially until each scenic spot successfully extracts 20 high-frequency feature words. Four scenic spots were removed due to the insufficient number of comments and the inaccurate extraction of high-frequency words.

I sort out the vocabulary of the high-frequency feature words in the network review data text of the scenic spot,

FIGURE 8: Proportion of positive emotions in the network data of rural tourist attractions.

so as to extract the common characteristics of the tourist demand of the scenic spot. There are 99 common characteristic words in rural tourism experience, as shown in Figure 7. The characteristics of rural tourism tourist experience in a certain area and the overall characteristics of tourist demand revealed from it are basically consistent with the characteristics of rural tourism demand. It is also explained from the side that the text analysis method used to evaluate the demand attributes of rural tourism reflects high accuracy and reliability.

5.5. Market Satisfaction and Overall Evaluation. I analyze tourists' overall satisfaction with the scenic spot based on the proportion of positive emotions in text documents and further understand the structural gap between industry supply and demand. In order to comprehensively weigh the overall satisfaction of tourists with rural tourism products in a certain area, first, the sentiment analysis method is used to calculate the proportion of positive emotions and negative emotions in the network text data of scenic spots. The positive emotions in most of the tourist spots' online review data texts are more than 70%, indicating that tourists' overall satisfaction with the rural tourism experience is relatively high. Considering the small number of words in a single online review data, about 30% of the negative sentiment proportion in the online review data also indicates that there is still room for improvement in the product quality and service quality of rural tourism in a certain area. Due to the large differences in the amount of network data information in different scenic spots, in order to more objectively reflect the differentiated characteristics of different scenic spots, the proportions of positive emotions in Internet word-of-mouth with different feature words are compared here. As shown in Figure 8, the FP_Apriori algorithm performs better than the Apriori algorithm in the proportion of positive emotions.

6. Conclusion

This paper improves an improved Apriori algorithm based on the frequent pattern tree FP-tree. The idea of this improved algorithm is to combine the traditional Apriori algorithm with FP-tree, which can operate on transactions, compress them efficiently, and avoid a large number of repeated transaction database traversals and scans. A new transaction database scanning strategy is proposed, which can scan only specific databases, making the scanning more purposeful and pertinent, and avoiding a lot of useless operations. The empirical research concludes that basic tourism consumption such as catering is dominant as a rural tourism demand motive, and demand motives such as leisure vacation and in-depth experience are insufficient. Tourists describe the experience content more than their spiritual perception, reflecting the slower and slower demand escalation. The overall satisfaction of rural tourists is relatively high, but they are also more dissatisfied with the level of service, the level of reception facilities, and tourism information services. Generally speaking, there are still big problems in the visit rate, popularity, and reputation of most rural tourist destinations in a certain area. In addition, the use of text analysis methods to extract rural tourism needs and experience characteristics has high accuracy. The travel motives, consumption characteristics, and organizational characteristics of rural tourism demand in a certain area extracted by the text analysis method are basically consistent with the rural tourism demand characteristics summarized in the literature review. The above empirical analysis reveals the structural contradictions within and between the two main bodies of rural tourism supply and demand. The supply of rural tourism urgently needs to innovate products and services to provide tourists with rural tourism experience products that match the quality and price and can lead the trend. The optimization of the rural tourism market structure also

requires the establishment of a benign interaction mechanism between industrial supply and industrial demand, leading demand upgrading through supply innovation, and promoting supply renewal through demand upgrading, forming a continuous driving force for the overall progress of the industry.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author does not have any possible conflicts of interest.

References

- [1] K. Pitchayadejanant and P. Nakpathom, "Data mining approach for arranging and clustering the agro-tourism activities in orchard," *Kasetsart Journal of Social Sciences*, vol. 39, no. 3, pp. 407–413, 2018.
- [2] F. Jiang, K. K. R. Yuen, and E. W. M. Lee, "Analysis of motor-cycle accidents using association rule mining-based framework with parameter optimization and GIS technology," *Journal of Safety Research*, vol. 75, pp. 292–309, 2020.
- [3] I. Khatri, "Information technology in tourism & hospitality industry: a review of ten years publications," *Journal of Tourism and Hospitality Education*, vol. 9, pp. 74–87, 2019.
- [4] Y. Li, H. Zhang, D. Zhang, and R. Abrahams, "Mediating urban transition through rural tourism," *Annals of Tourism Research*, vol. 75, pp. 152–164, 2019.
- [5] N. Cucari, E. Wankowicz, and S. E. De Falco, "Rural tourism and Albergo Diffuso: a case study for sustainable land-use planning," *Land Use Policy*, vol. 82, pp. 105–119, 2019.
- [6] W. Yu, "Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 855–866, 2019.
- [7] F. Xu, N. Nash, and L. Whitmarsh, "Big data or small data? A methodological review of sustainable tourism," *Journal of Sustainable Tourism*, vol. 28, no. 2, pp. 144–163, 2020.
- [8] P. Siakwah, R. Musavengane, and L. Leonard, "Tourism governance and attainment of the sustainable development goals in Africa," *Tourism Planning & Development*, vol. 17, no. 4, pp. 355–383, 2020.
- [9] M. M. Su, G. Wall, and S. Wang, "Yujiale fishing tourism and island development in Changshan archipelago, Changdao, China," *Island Studies Journal*, vol. 12, no. 2, pp. 127–142, 2017.
- [10] Q. He, W. He, Y. Song, J. Wu, C. Yin, and Y. Mou, "The impact of urban growth patterns on urban vitality in newly built-up areas based on an association rules analysis using geographical 'big data'," *Land Use Policy*, vol. 78, pp. 726–738, 2018.
- [11] H. Q. Vu, G. Li, R. Law, and Y. Zhang, "Tourist activity analysis by leveraging mobile social media data," *Journal of Travel Research*, vol. 57, no. 7, pp. 883–898, 2018.
- [12] W. Wang, J. Wu, M. Y. Wu, and P. L. Pearce, "Shaping tourists' green behavior: the hosts' efforts at rural Chinese B&Bs," *Journal of Destination Marketing & Management*, vol. 9, pp. 194–203, 2018.
- [13] R. Manikandan and V. Saravanan, "A novel approach on particle agent swarm optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective," *Multimedia Tools and Applications*, vol. 79, no. 5–6, pp. 3807–3829, 2020.
- [14] I. O. Ezeuduji, "Change management for sub-Saharan Africa's rural tourism development," *Current Issues in Tourism*, vol. 20, no. 9, pp. 946–959, 2017.
- [15] C. H. Chin, M. C. Lo, A. A. Mohamad, and V. Nair, "The impacts of multi-environmental constructs on tourism destination competitiveness: local residents' perceptions," *Journal of Sustainable Development*, vol. 10, no. 3, pp. 120–132, 2017.
- [16] F. Wang, L. Xia, Z. Chen, W. Cui, Z. Liu, and C. Pan, "Remote sensing identification of coastal zone mariculture modes based on association-rules object-oriented method," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, no. 12, pp. 210–217, 2018.
- [17] X. Zhao, X. Lu, Y. Liu, J. Lin, and J. An, "Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: a case study of Xi'an, China," *Tourism Management*, vol. 69, pp. 368–383, 2018.
- [18] J. Yuan, H. Lin, and X. Li, "Analysis and comments on the course of rural development," *Agricultural Science & Technology*, vol. 18, no. 12, pp. 2653–2657, 2017.
- [19] K. Güler and Y. Kâhya, "Developing an approach for conservation of abandoned rural settlements in Turkey," *A/Z ITU Journal of the Faculty of Architecture*, vol. 16, no. 1, pp. 97–115, 2019.
- [20] T. C. Leu, M. Eriksson, and D. K. Müller, "More than just a job: exploring the meanings of tourism work among indigenous Sámi tourist entrepreneurs," *Journal of Sustainable Tourism*, vol. 26, no. 8, pp. 1468–1482, 2018.
- [21] X. Yang and P. Ho, "Mining institutions, contention and credibility: applying the conflict analysis model to court cases in China," *The Extractive Industries and Society*, vol. 7, no. 3, pp. 1011–1021, 2020.
- [22] R. Armis and H. Kanegae, "The attractiveness of a post-mining city as a tourist destination from the perspective of visitors: a study of Sawahlunto old coal mining town in Indonesia," *Asia-Pacific Journal of Regional Science*, vol. 4, no. 2, pp. 443–461, 2020.